



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 29, 2020 – 08:36 am BST

PDB ID : 5VH2  
Title : Crystal Structure of Mouse Cadherin-23 EC12-13 with Engineered Mutation S1339D  
Authors : Termine, D.J.; Jaiganesh, A.; Sotomayor, M.  
Deposited on : 2017-04-12  
Resolution : 2.84 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

---

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

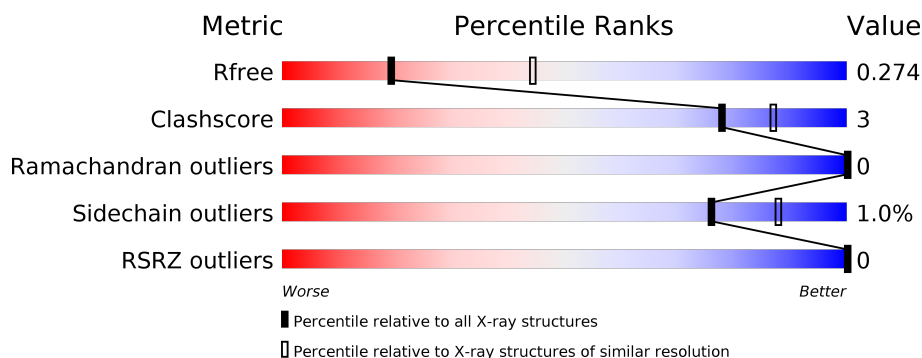
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.84 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	1031 (2.86-2.82)
Clashscore	141614	1078 (2.86-2.82)
Ramachandran outliers	138981	1050 (2.86-2.82)
Sidechain outliers	138945	1051 (2.86-2.82)
RSRZ outliers	127900	1019 (2.86-2.82)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions  $\leq 5\%$ . The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	220	 88% 7% 5%
1	B	220	 89% 6% 5%
1	C	220	 86% 6% 8%
1	D	220	 83% 11% 5%

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 6390 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Cadherin-23.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	209	Total	C	N	O	S	0	0	0
			1606	1015	257	331	3			
1	B	208	Total	C	N	O	S	0	0	0
			1598	1011	256	328	3			
1	C	203	Total	C	N	O	S	0	0	0
			1572	996	250	323	3			
1	D	208	Total	C	N	O	S	0	0	0
			1600	1012	255	330	3			

There are 40 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1178	MET	-	initiating methionine	UNP Q99PF4
A	1316	ASP	SER	engineered mutation	UNP Q99PF4
A	1390	LEU	-	expression tag	UNP Q99PF4
A	1391	GLU	-	expression tag	UNP Q99PF4
A	1392	HIS	-	expression tag	UNP Q99PF4
A	1393	HIS	-	expression tag	UNP Q99PF4
A	1394	HIS	-	expression tag	UNP Q99PF4
A	1395	HIS	-	expression tag	UNP Q99PF4
A	1396	HIS	-	expression tag	UNP Q99PF4
A	1397	HIS	-	expression tag	UNP Q99PF4
B	1178	MET	-	initiating methionine	UNP Q99PF4
B	1316	ASP	SER	engineered mutation	UNP Q99PF4
B	1390	LEU	-	expression tag	UNP Q99PF4
B	1391	GLU	-	expression tag	UNP Q99PF4
B	1392	HIS	-	expression tag	UNP Q99PF4
B	1393	HIS	-	expression tag	UNP Q99PF4
B	1394	HIS	-	expression tag	UNP Q99PF4
B	1395	HIS	-	expression tag	UNP Q99PF4
B	1396	HIS	-	expression tag	UNP Q99PF4
B	1397	HIS	-	expression tag	UNP Q99PF4
C	1178	MET	-	initiating methionine	UNP Q99PF4

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
C	1316	ASP	SER	engineered mutation	UNP Q99PF4
C	1390	LEU	-	expression tag	UNP Q99PF4
C	1391	GLU	-	expression tag	UNP Q99PF4
C	1392	HIS	-	expression tag	UNP Q99PF4
C	1393	HIS	-	expression tag	UNP Q99PF4
C	1394	HIS	-	expression tag	UNP Q99PF4
C	1395	HIS	-	expression tag	UNP Q99PF4
C	1396	HIS	-	expression tag	UNP Q99PF4
C	1397	HIS	-	expression tag	UNP Q99PF4
D	1178	MET	-	initiating methionine	UNP Q99PF4
D	1316	ASP	SER	engineered mutation	UNP Q99PF4
D	1390	LEU	-	expression tag	UNP Q99PF4
D	1391	GLU	-	expression tag	UNP Q99PF4
D	1392	HIS	-	expression tag	UNP Q99PF4
D	1393	HIS	-	expression tag	UNP Q99PF4
D	1394	HIS	-	expression tag	UNP Q99PF4
D	1395	HIS	-	expression tag	UNP Q99PF4
D	1396	HIS	-	expression tag	UNP Q99PF4
D	1397	HIS	-	expression tag	UNP Q99PF4

- Molecule 2 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	B	3	Total Ca 3 3	0	0
2	A	3	Total Ca 3 3	0	0
2	D	2	Total Ca 2 2	0	0
2	C	2	Total Ca 2 2	0	0

- Molecule 3 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	1	Total Na 1 1	0	0
3	A	1	Total Na 1 1	0	0
3	D	1	Total Na 1 1	0	0

*Continued on next page...*


*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	C	1	Total	Na	0	0
			1	1		

### 3 Residue-property plots [i](#)


These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ( $RSRZ > 2$ ). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

#### • Molecule 1: Cadherin-23

Chain A: 




#### • Molecule 1: Cadherin-23

Chain B: 




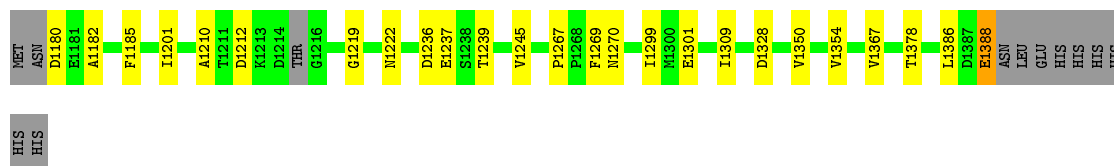
#### • Molecule 1: Cadherin-23

Chain C: 



#### • Molecule 1: Cadherin-23

Chain D: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	43.74Å 47.21Å 113.04Å 89.95° 90.02° 115.76°	Depositor
Resolution (Å)	33.99 – 2.84 33.99 – 2.82	Depositor EDS
% Data completeness (in resolution range)	92.7 (33.99-2.84) 92.8 (33.99-2.82)	Depositor EDS
$R_{merge}$	0.17	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.15 (at 2.81Å)	Xtriage
Refinement program	REFMAC 5.8.0158	Depositor
R, $R_{free}$	0.238 , 0.272 0.242 , 0.274	Depositor DCC
$R_{free}$ test set	1084 reflections (5.96%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	10.8	Xtriage
Anisotropy	1.327	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , -12.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	0.043 for h,-h-k,-l 0.370 for -h,-k,l 0.036 for -h,h+k,-l	Xtriage
Reported twinning fraction	0.564 for H, K, L 0.436 for -h,-k,l	Depositor
Outliers	0 of 18188 reflections	Xtriage
$F_o, F_c$ correlation	0.83	EDS
Total number of atoms	6390	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	24.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.25% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup> Intensities estimated from amplitudes.

<sup>2</sup> Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

## 5 Model quality

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: NA, CA

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 5$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	$\# Z  > 5$	RMSZ	$\# Z  > 5$
1	A	0.58	0/1629	0.79	1/2217 (0.0%)
1	B	0.60	0/1621	0.79	1/2206 (0.0%)
1	C	0.62	0/1594	0.84	3/2168 (0.1%)
1	D	0.58	0/1622	0.78	0/2205
All	All	0.59	0/6466	0.80	5/8796 (0.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1
1	B	0	1
1	C	0	1
1	D	0	1
All	All	0	4

There are no bond length outliers.

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	1303	LEU	CA-CB-CG	7.42	132.37	115.30
1	C	1237	GLU	CB-CA-C	-7.28	95.84	110.40
1	C	1214	ASP	CB-CG-OD2	5.84	123.55	118.30
1	A	1355	ASP	CB-CG-OD1	5.11	122.89	118.30
1	B	1237	GLU	CB-CA-C	-5.09	100.23	110.40

There are no chirality outliers.

All (4) planarity outliers are listed below:



Mol	Chain	Res	Type	Group
1	A	1245	VAL	Peptide
1	B	1245	VAL	Peptide
1	C	1245	VAL	Peptide
1	D	1245	VAL	Peptide

## 5.2 Too-close contacts

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1606	0	1573	10	0
1	B	1598	0	1569	7	0
1	C	1572	0	1541	8	0
1	D	1600	0	1565	17	0
2	A	3	0	0	0	0
2	B	3	0	0	0	0
2	C	2	0	0	0	0
2	D	2	0	0	0	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
3	C	1	0	0	0	0
3	D	1	0	0	0	0
All	All	6390	0	6248	37	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 3.

The worst 5 of 37 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:1360:ASP:OD2	1:B:1215:THR:OG1	2.14	0.65
1:A:1309:ILE:HD11	1:A:1350:VAL:HG23	1.82	0.61
1:C:1303:LEU:HG	1:C:1350:VAL:HG21	1.83	0.60
1:B:1299:ILE:HG21	1:B:1309:ILE:HD12	1.81	0.60
1:A:1326:ARG:HE	1:D:1201:ILE:HG21	1.70	0.57

There are no symmetry-related clashes.

## 5.3 Torsion angles [i](#)

### 5.3.1 Protein backbone [i](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	207/220 (94%)	201 (97%)	6 (3%)	0	100	100
1	B	206/220 (94%)	197 (96%)	9 (4%)	0	100	100
1	C	199/220 (90%)	189 (95%)	10 (5%)	0	100	100
1	D	204/220 (93%)	192 (94%)	12 (6%)	0	100	100
All	All	816/880 (93%)	779 (96%)	37 (4%)	0	100	100

There are no Ramachandran outliers to report.

### 5.3.2 Protein sidechains [i](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	175/186 (94%)	174 (99%)	1 (1%)	86	93
1	B	174/186 (94%)	173 (99%)	1 (1%)	86	93
1	C	172/186 (92%)	170 (99%)	2 (1%)	71	85
1	D	174/186 (94%)	171 (98%)	3 (2%)	60	80
All	All	695/744 (93%)	688 (99%)	7 (1%)	76	88

5 of 7 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	1328	ASP
1	D	1388	GLU
1	D	1180	ASP

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	1238	SER
1	D	1328	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	1270	ASN

### 5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

### 5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

### 5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

### 5.6 Ligand geometry [i](#)

Of 14 ligands modelled in this entry, 14 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

### 5.7 Other polymers [i](#)

There are no such residues in this entry.

## 5.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.

## 6 Fit of model and data ⓘ

### 6.1 Protein, DNA and RNA chains ⓘ

In the following table, the column labelled ‘#RSRZ> 2’ contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95<sup>th</sup> percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled ‘Q< 0.9’ lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å <sup>2</sup> )	Q<0.9
1	A	209/220 (95%)	-0.61	0 100 100	18, 25, 34, 43	0
1	B	208/220 (94%)	-0.65	0 100 100	17, 22, 39, 48	0
1	C	203/220 (92%)	-0.77	0 100 100	13, 19, 30, 35	0
1	D	208/220 (94%)	-0.61	0 100 100	21, 27, 35, 46	0
All	All	828/880 (94%)	-0.66	0 100 100	13, 24, 35, 48	0

There are no RSRZ outliers to report.

### 6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

### 6.3 Carbohydrates ⓘ

There are no carbohydrates in this entry.

### 6.4 Ligands ⓘ

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled ‘Q< 0.9’ lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å <sup>2</sup> )	Q<0.9
2	CA	D	1402	1/1	0.96	0.07	23,23,23,23	0
2	CA	B	1401	1/1	0.96	0.09	21,21,21,21	0
3	NA	D	1403	1/1	0.96	0.08	28,28,28,28	0
3	NA	B	1404	1/1	0.98	0.08	26,26,26,26	0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
3	NA	C	1403	1/1	0.98	0.08	16,16,16,16	0
2	CA	A	1401	1/1	0.98	0.07	26,26,26,26	0
2	CA	A	1403	1/1	0.99	0.04	27,27,27,27	0
2	CA	D	1401	1/1	0.99	0.06	22,22,22,22	0
2	CA	B	1402	1/1	0.99	0.09	19,19,19,19	0
3	NA	A	1404	1/1	0.99	0.10	28,28,28,28	0
2	CA	C	1402	1/1	0.99	0.10	16,16,16,16	0
2	CA	A	1402	1/1	0.99	0.07	27,27,27,27	0
2	CA	B	1403	1/1	0.99	0.09	19,19,19,19	0
2	CA	C	1401	1/1	0.99	0.07	16,16,16,16	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.